Mapping local and global variability in plant trait distributions

Ethan E. Butler[1,a,b], Abhirup Datta[2,a,b], Habacuc Flores-Moreno[1,3], Ming Chen[1], Kirk R. Wythers[1], Farideh Fazayeli[4], Arindam Banerjee[4], Owen K. Atkin[5,6], Jens Kattge[7,8], Bernard Amiaud[9], Benjamin Blonder[10], Gerhard Boenisch[7], Ben Bond-Lamberty[11], Kerry A. Brown[12], Chaeho Byun[13], Giandiego Campetella[14], Bruno E.L. Cerabolini[15], Johannes H.C. Cornelissen[16], Joseph M. Craine[17], Dylan Craven[8,18], Franciska T. de Vries[19], Sandra Díaz[20], Tomas Domingues[21], Estelle Forey[22], Andres Gonzalez[23], Nicolas Gross[24,25,26], Wenxuan Han[27,28], Wesley N. Hattingh[29], Thomas Hickler[30,31], Steven Jansen[32], Koen Kramer[33], Nathan J.B. Kraft[34], Hiroko Kurokawa[35], Daniel C. Laughlin[36], Patrick Meir[6,37], Vanessa Minden[38], Ülo Niinemets[29], Yusuke Onoda[40], Josep Peñuelas[41,42], Quentin Read[43], Fernando Valladares Ros[44], Lawren Sack[34], Brandon Schamp[45], Nadejda A. Soudzilovskaia[46], Marko J. Spasojevic[47], Enio Sosinski[48], Peter Thornton[49], Peter M. van Bodegom[46], Mathew Williams[37], Christian Wirth[7,8,50], and Peter B. Reich[1,51]

[1]Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; [2]Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21202; [3]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, 55108; [4]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, 55455 ; [5]ARC Centre of Excellence in Plant Energy, Research School of Biology, The Australian National University, Building 134, Canberra, ACT 2601, Australia; [6]Division of Plant Sciences, Research School of Biology, the Australian National University, Building 134, Canberra, ACT 2601, Australia; [7]Max Planck Institute for Biogeochemistry, Hans Knoell Strasse 10, 07745, Jena, Germany; [8]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany; [9]UMR 1137 Ecologie et Ecophysiologie Forestières, Université de Lorraine – INRA, 54506 Vandoeuvre-lès-Nancy, France; [10]Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3BJ, United Kingdom; [11]Joint Global Change Research Institute, DOE Pacific Northwest National Laboratory, College Park, MD USA; [12]Department of Geography and Geology, School of Natural and Built Environments, Kingston University London, Penrhyn Road, Surrey, KT1 2EE; [13]School of Biological Sciences, Seoul National University, Seoul 08826, South Korea; [14]School of Biosciences & Veterinary Medicine, Plant Diversity and Ecosystems Management unit, University of Camerino, Italy; [15]Department of Theoretical and Applied Sciences, University of Insubria, Via J.H. Dunant 3, I-21100 Varese, Italy; [16]Systems Ecology, Department of Ecological Science, Vrije Universiteit, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; [17]Jonah Ventures, Manhattan KS 66502; [18]Department of Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser Straße 4, 06120, Halle (Saale), Germany; [19]School of Earth and Environmental Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom; [20]Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET) and Departamento de Diversidad Biológica y Ecología, FCEFyN, Universidad Nacional de Córdoba, CC 495, Córdoba, Argentina; [21]Faculdade de Filosofia Ciencias e Letras de Ribeirao Preto, Universidade de Sao Paulo, Av Bandeirantes, 3900, CEP 14040-901, Bairro Monte Alegre, Ribeirao Preto, Sao Paulo, Brazil; [22]Normandie University, UNIROUEN, IRSTEA, ECODIV, FR-76000 Rouen, France; [23]Universidad del Rosario. Facultad de Ciencias Naturales y Matematicas. Carrera 26 No 63B-48, Bogota, Colombia; [24]Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, C/ Tulipán s/n, 28933 Móstoles, Spain; [25]INRA, USC1339 Chizé (CEBC), F-79360, Villiers en Bois, France; [26]Centre d'étude biologique de Chizé, CNRS - Université La Rochelle (UMR 7372), F-79360, Villiers en Bois, France; [27]College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China; [28]Key Laboratory of Biogeography and Bio-resource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; [29]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, WITS 2050, Johannesburg, South Africa; [30]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt/Main, Germany; [31]Department of Physical Geography at Goethe-University, Frankfurt/Main; [32]Ulm University, Institute of Systematic Botany and Ecology, Albert-Einstein-Allee 11, 89081 Ulm, Germany; [33]Wageningen Environmental Research (Alterra); [34]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, 90095; [35]Forestry and Forest Products Research Institute 1 Matsunosato, Tsukuba 305-8687, Japan; [36]Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, US; [37]School of Geosciences, University of Edinburgh, Edinburgh, EH9 3FF, UK; [38]Institute of Biology and Environmental Science, University of Oldenburg, Carl von Ossietzky-Straße 9-11, 26111, Oldenburg, Germany; [39]Department of Plant Physiology, Estonian University of Life Sciences, Kreutzwaldi 1, 51014 Tartu, Estonia; [40]Graduate School of Agriculture, Kyoto University, Kyoto, 606-8502 Japan; [41]CSIC, Unitat d'Ecologia Global CREAF-CSIC-UAB, Bellaterra 08193, Barcelona, Catalonia, Spain; [42]CREAF, Cerdanyola del Vallès 08193, Barcelona, Catalonia, Spain; [43]Michigan State University, Department of Forestry, 480 Wilson Rd., East Lansing, MI 48824; [44]Museo Nacional de Ciencias Naturales, CSIC Serrano 115 dpdo, E-28006 Madrid Spain; [45]Dept. of Biology, Algoma University, Sault Ste. Marie, Ontario, Canada P6A 2G4; [46]Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands; [47]Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, Riverside, CA. 92521; [48]Embrapa Clima Temperado, Rodovia BR 392, Km 78, Pelotas, RS, Brasil, 96010-971; [49]Environmental Sciences Division, Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA; [50]Department Systematic Botany and Functional Biodiversity, University of Leipzig, 04103 Leipzig, Germany; [51]Hawkesbury Institute for the Environment, Western Sydney University, Penrith New South Wales 2751, Australia

Our ability to understand and predict the response of ecosystems to a changing environment depends on quantifying vegetation functional diversity. However, representing this diversity at the global scale is challenging. Typically, in Earth Systems Models, characterization of plant diversity has been limited to grouping related species into Plant Functional Types (PFTs), with all trait variation in a PFT collapsed into a single mean value that is applied globally. Using the largest global plant trait database and state of the art Bayesian modeling, we created fine-grained global maps of plant trait distributions that can be applied to Earth System Models. Focusing on a set of plant traits closely coupled to photosynthesis and foliar respiration – specific leaf area (SLA), and dry mass-based concentrations of leaf nitrogen (N_m) and phosphorus (P_m), we characterize how traits vary within and among over 50,000 ≈50 km cells across the entire vegetated land surface. We do this in several ways - without defining the PFT of each grid cell, and using 4 or 14 PFTs; each model's predictions are evaluated against out-of-sample data. This endeavor advances prior trait mapping by generating global maps that preserve variability across scales by using modern Bayesian spatial statistical modeling in combination with a database over three times larger than previous analyses (Van Bodegom, et al. 2014; Maire, et al., 2015). Our maps reveal that the most diverse pixels possess trait variability close to the range of global PFT means.

plant traits | Bayesian modeling | spatial statistics | global | climate

Modeling global climate and the carbon cycle with Earth System Models (ESMs) requires maps of plant traits that play key roles in leaf- and ecosystem-level metabolic processes (1–4). Multiple traits are critical to both photosynthesis and respiration, foremost leaf nitrogen concentration (N_m) and specific leaf area (SLA) (5–7). More recently, variation in leaf phosphorus concentration (P_m) has also been linked to variation in photosynthesis and foliar respiration (7–12). Estimating detailed global geographic patterns of these traits and corresponding trait-environment relationships has been hampered by limited measurements (13), but recent improvements in data coverage (14) allows for greater detail in spatial estimates of these key traits.

Previous work has extrapolated trait measurements across continental or larger regions through three methodologies: grouping measurements of individuals into larger categories that share a set of properties (a working definition of plant functional types or PFTs) (4, 15), exploiting trait-environment

relationships (e.g. leaf N_m and mean annual temperature) (1, 16–19), or restricting the analysis to species whose presence has been widely estimated on the ground (20–23). Each of these methods has limitations - for example, trait-environment relationships do not well explain observed trait spatial patterns(1, 24), while species-based approaches limit the scope of extrapolation to only areas with well measured species abundance. More critically, the first two global methodologies emphasized estimating a single trait value per PFT at every location, whereas both ground based (5, 14) and remotely sensed (25) observations suggest that at ecosystem or landscape scales traits would be better represented by distributions. Here, we use an updated version of the largest global database of plant traits (14) coupled with modern Bayesian spatial statistical modeling techniques (26) to capture local and global variability in plant traits. This combination allows the representation of trait variation both within pixels on a gridded land surface as well as across global environmental gradients.

Information is lost when the range of measured trait values is compressed into a single PFT (Fig. 1). We observe that the global range of site level SLA values for a single PFT such as Broadleaf Evergreen Tropical trees (Fig. 1a,c) is quite large (2.7 to 65.2 $m^2\ kg^{-1}$). Even after limiting the scope to a single well measured $0.5^\circ \times 0.5^\circ$ pixel within Panama (Fig. 1b,c), there is still a wide range of SLA values (4.7 to 37.7 $m^2\ kg^{-1}$) with a local mean of 15.7 $m^2\ kg^{-1}$, and a local standard deviation of 5.4 $m^2\ kg^{-1}$ – over 1/3 of the local mean. By contrast, the mean SLA value of all species associated with Broadleaf Evergreen Tropical trees is 13.9 $m^2\ kg^{-1}$, over 10% lower than the local average (Fig. 1c). Thus, single trait values per PFT fail to capture variability in trait values within or among grid cells; i.e. over a wide range of spatial scales.

Transitioning from a single trait value per PFT to a distribution may lead to significantly different modeling results as critical plant processes, such as photosynthesis, are non-linear with respect to these traits (27). This is reinforced by recent modeling studies which have begun to incorporate distributions of traits at regional (28, 29) and global (30) scales. It has been shown that using trait distributions leads to different estimates of carbon dynamics (31) and that higher order moments of trait distributions contribute to sustaining multiple ecosystem functions (32). While species level mapping (20, 22, 23) does capture trait distributions, it has been limited geographically and restricted to subsets of functional groups.

Even the largest plant trait database offers only partial coverage across the globe in terms of site level measurements. Hence, gap-filling approaches need to be adopted to extrapolate trait values at regions with no data coverage. Here, we overcome data limitations through PFT classification, trait-environment relationships, and additional location information to develop a suite of models capable of estimating trait distributions across the entire vegetated globe. The simplest is a categorical model, which assigns traits to maps of remotely sensed PFTs. Every species, with its corresponding trait values, is associated with a PFT and these trait distributions are extrapolated to the satellite estimated range of the PFT (Figs. S1-S2). The second is a Bayesian linear model which complements the PFT information with trait-environment relationships. The third is a Bayesian spatial model which, in addition to PFTs and the trait-environment relationships, leverages additional location information via Gaussian Processes (Methods). The use of a spatial Gaussian Process in this context is novel and model evaluation reveals the superior predictive performance of this model.

Each of these methods interpolate (and extrapolate) both mean trait values and entire trait distributions across space (i.e. across grid cells on a global map). These models are further stratified by three different levels of PFT categorization: 1) global, all plants in a single group; 2) broad, four groups based on growth form and leaf type; 3) narrow, fourteen groups based on further environmental, phenological, and photosynthetic categories (Methods). The global categorization groups all plants into a single class, while the broad grouping (4-PFT) is similar to the vegetation classification used in the JULES land surface (33), and the narrow (14-PFT) categories are equivalent to the classification used in the Community Land Model (4, 15, 34).

The above mentioned methods allow for a representation of global vegetation that enables a more accurate formulation of functional diversity than the single trait value per PFT paradigm that is widely employed (4). The traits studied here - SLA, N_m, and P_m - are central to predicting variation in rates of plant photosynthesis (5, 6, 9, 11) and foliar respiration (10, 35). The importance of these traits and the more advanced representation of functional diversity developed here may be used to better capture the response of the land surface component of the Earth System to environmental change.

Significance Statement

Currently, Earth System Models (ESMs) represent variation in plant life through the presence of a small set of Plant Functional Types (PFTs), each of which accounts for hundreds or thousands of species across thousands of vegetated grid cells on land. By expanding plant traits from a single mean value per PFT to a full distribution per PFT that varies among grid cells, the trait variation present in nature is restored and may be propagated to estimates of ecosystem processes. Indeed, critical ecosystem processes tend to depend on the full trait distribution, which therefore needs to be represented accurately. These maps re-introduce substantial local variation and will allow for a more accurate representation of the land surface in ESMs.

The general idea for the study was developed by E.E.B., A.D, H.F.M., F.F., M.C., K.W., A.B., J.K., O.K.A., and P.B.R.; specifics were developed by E.E.B. and A.D., and refined with the rest of that team. Data were made available by the hundreds of contributors to the TRY database; with further data management and compilation by E.E.B, A.D., H.F.M. and J.K. E.E.B. and A.D. performed the analysis, with all authors contributing to interpretation. E.E.B. and A.D. wrote the first draft; all authors contributed to subsequent versions, including the submitted one.

The authors declare no known conflicts of interest.

[a]E.E.B and A.D. contributed equally to this work.

[b]To whom correspondence should be addressed. E-mail: eebutler@umn.edu, abhidatta@jhu.edu

Results and Discussion

Model Evaluation. Given the full suite of nine models proposed, we conducted extensive model evaluation (see Table 1) to determine the trade-offs associated with each methodology and resolution of PFT. We assessed the predictive capability of the models using the root mean squared predictive error (RMSPE) based on out-of-sample data (see Section S6). Among the nine models, the spatial narrow 14-PFT model emerged as

DRAFT

the best predictor of mean trait values for SLA and N_m, and the second best for P_m (Table 1). However, the spatial broad 4-PFT model performed nearly as well (Table 1). The close predictive skill of the broad PFT model and the narrow PFT model suggests that variations in growth form and leaf type, used to define the 4 broad PFTs, are more strongly related to variations of trait values than bioclimatic differences, which are used in the global spatial model. The models' abilities to correctly estimate the spread of the trait distributions were assessed using the out-of-sample coverage probabilities (CP) – the proportion of instances the model predicted 95% confidence intervals contained the observed trait values. Most of the models provided adequate coverage (CP of around 90% or more). See Section S4 for more detailed definitions of the model comparison metrics.

The improvement in prediction afforded by the inclusion of a spatial term and PFT information invites further evaluation (Table 1). Spatial relationships play a substantial role in shaping trait values, at least at the global scale. The spatial term in the model is likely to approximate finer scale variation that is unavailable for the large grid cells of global analyses. For example, smaller scale studies that can evaluate local variations in climate, soil, or other relevant abiotic or biotic covariates are unlikely to see as much improvement with the use of a spatial term, as they may directly measure local sources of variability. However, with large grid cells the spatial term allows for local adjustment of trait values that the relatively coarse environmental metrics are not able to capture. Similarly, the use of PFTs greatly improves the models. The greatest decrease in RMSPE occurs between the global grouping (single global PFT-equivalent) and the broad (4-PFT) grouping across each of the models tested. This implies that the broad PFTs, based primarily on growth form and leaf type, offer superior predictive skill than environmental covariates on their own (19). However, the extra information in the narrow (14-PFT) grouping does further improve the fit and produces the most accurate predicted trait surface.

Global Maps. As the narrow 14-PFT spatial model is the best predictor of mean trait values, and provided adequate coverage probability, we show the mean and standard deviation maps for this model (Fig. 2-4a,b). To provide a comparison, we also included the best map closest to those currently employed in ESMs - the narrow PFT categorical model (Fig. 2-4c,d). Maps for the other models can be found in the supplemental material (Figs. S8-S16). The mean and standard deviation are presented as a summary of the full log-normal distribution within each pixel, but there are full distributions estimated in each pixel, see Case Studies below.

The standard deviation maps (Figs. 2-4b,d) compared to the mean maps (Figs. 2-4a,c) highlight one of the central results of this analysis – the local standard deviations of trait values are of similar magnitudes as their respective means. Generally, we observed that the local standard deviation is close to half the local mean value but can approach the global range of the trait mean values, e.g. N_m (Fig. 3) has a maximum local standard deviation of 9 mg N / g, and the global mean range is only ≈10 mg N / g. The maps of the trait standard deviations follow similar patterns to the means, though there are several regions where the mean varies more rapidly than the standard deviation; such as SLA in the SE United States and China in the categorical model (Fig. 2c,d) and similarly for N_m in the spatial model across the Sahel in sub-Saharan Africa (3a,c). The lack of variation in the standard deviation is most clear in the categorical model for N_m while both models show relatively modest variation in P_m.

While the broad features of the mean maps for both the categorical and spatial models are similar across SLA (Fig. 2), N_m (Fig. 3), and P_m (Fig. 4); there are numerous marked differences across both broad and fine spatial scales. The shared broad features of the maps from both models include SLA and P_m increasing from the tropics to the poles, while N_m has more modest variation, except that it tends to be lower in regions dominated by needle-leaved trees. Some of the notable differences between the models include the spatial model's greater range and more marked variability of SLA within equatorial regimes (e.g., Brazil or central Africa); it also better captures the low SLA of most of arid Australia than the categorical model; and more strongly highlights the gradient of P_m from the tropics to the arctic (16) (Fig 2a,4a).

The most consistent estimates between the categorical and spatial models are in the boreal regions dominated by needle-leaf trees; the measurements in this region are relatively sparse which may have limited the ability of the spatial model to capture differences. On the other hand, broad-leaved trees span a wide range of environments, but a large portion of the measurements come from the tropics (66%), where there is a limited range of values among the climate covariates and therefore little variation with which to estimate a correlation. The grasses and shrubs have the largest standard deviations of the four broad PFTs (Table S4) and dominate wide swathes of the land surface, but have fewer measurements – shrubs are the least measured of the broad PFTs in the database, and this appears to reduce the accuracy of the categorical model more than the spatial model (Table 1). The fact that shrubs are assumed to dominate in arid and boreal environments, which also tend to be under-sampled, also likely contributes to these differences.

Our results suggest that the breadth of functional niche space is reduced in both boreal and tropical biogeographic regions. The low variation across all three traits within the boreal forest implies that there is strong filtering and smaller niche space available in this relatively harsh environment. Surprisingly, despite the high species diversity in tropical forests, we also find that SLA and P_m have relatively low variation in these forests – suggesting that in this environment the trait space is reduced. This could be, in part, an artifact of the Earth System Model PFT classification omitting herbaceous species. Conversely, grasslands and savannahs exhibit large variation in total trait space, suggesting these environments permit a wider range of strategies than in both the boreal and tropical regions. Most broadly, both the data and the spatial model indicate (Figs. S24,S25) low leaf nitrogen values in temperate climates that increase in both colder and warmer regions; this may indicate a more complicated leaf biochemistry-temperature relationship than has previously been suggested (16).

Case Studies. As a more detailed evaluation of the true and predicted shape of the trait distributions cannot be obtained from the coverage probability or standard deviation maps, we conducted two regional case studies, in which trait data were pooled over an area to construct full trait distributions and these were formally compared with the model predicted

DRAFT

distributions.

We considered two areas with substantially different environmental conditions to evaluate the trait distributions obtained from the spatial and categorical models. We chose a single pixel that contained a highly studied site with numerous measurements of tropical trees, Barro Colorado Island (BCI), Panama, and a collection of pixels in an arid environment in which the mean estimates for SLA of the spatial and categorical models substantially disagreed, the southwestern United States. These areas were in the training data, and this analysis constituted a more detailed analysis of the models' fit to the observed distribution of these locations. Here, the focus was on the structure of the full distribution of traits predicted at these sites; Fig. S17 is a map of the measurements that comprised these locations and other sites included in this analysis. Both areas offer further insight into the structure of the distributions estimated by the categorical and spatial models.

In the pixel containing BCI, the categorical and spatial models broadly agreed for all three traits (Fig. 5a, c, e), although the spatial model means were only half as distant from the observed means for SLA and N_m (4% vs. 8% and 5% vs. 10%, respectively). There were only two PFTs present in this pixel: tropical broadleaf evergreen and deciduous trees. Despite the general similarity of the shapes of the distribution, the spatial model appears capable of capturing some subtle features. This is clearest for leaf nitrogen, where the peak of the distribution was quite broad. This is neatly captured in the narrow PFT model, and the pattern was detectable through the Kolmogorov-Smirnov (K-S) statistic, which evaluates the similarity of two full distributions. Indeed, the superiority of the spatial model was reinforced by a closer match for the Bayesian spatial model across all traits at BCI (Table S6), though for P_m it was the global spatial model that fit best (not shown in the figure).

The differences between the trait distributions of the categorical and Bayesian spatial models were stark in the southwestern United States, although the mean estimates for N_m and P_m were close (Fig. 5b, d, f). This may be a result of the topographic complexity of this region and the resulting difficulty of aggregating climate and soil covariates at the 0.5° pixel scale and the sparser sampling than at BCI. To get enough data to approximate a distribution, we aggregated 18 pixels with nine PFTs including every temperate category, though many of them are only marginally present. The inclusion of so many PFTs produced a noisier distribution in the categorical model than suggested by the data and estimated by the spatial model. Neither of the models produced distributions that matched as well with the observations; however, it is notable how close the mean values for both models matched the observations for N_m and P_m, and the spatial model did well for the mean SLA.

Environmental Covariates and the Spatial Term. The improvement in prediction from the linear model to the spatial model was partially explained by the weak trait-environment relationships presented in Tables S2-S4. The magnitude of the spatial variation explained with the Gaussian process model was of similar magnitude as the unexplained trait variation. For most of the spatial models, the estimated spatial range was around 500 kilometers; this suggests a strong spatial effect, and implies that the spatial model can provide more precise information about the trait distribution near the locations where we have data. This was largely borne out in the case studies, and is illustrated more explicitly in Fig. 6 where the predicted trait standard deviation for the spatial model was up to 50% lower than the linear non-spatial model near locations with trait measurements. The spatial model thus leverages local information to reduce the uncertainty of trait estimation near data locations and may provide guidance for future data collection by identifying high uncertainty regions.

Applications for Trait Distributions. Plant traits vary across a range of spatial scales, and the spatial model best captures the large spatial gradients (such as in Amazonia and Australia) as well as the subtleties within pixels. Maps for all the models highlight how much information about local variability is lost when representing plant traits with a single value, and suggests that a first application of these maps will be for ESMs to incorporate these scales of variability. For process-based ESMs, the simplest model to incorporate will likely be the categorical model as it is closest to the current PFT approach, but this model is also the least flexible. The more sophisticated models developed here provide more accurate large scale variation, and may be used to infer new trait values in a novel climate by perturbing the climatic covariates (36). However, given the likelihood of non-linear trait-environment relationships, the spatial sparsity of the data, and the possibility of alternate strategies within a PFT that may alter the trait-environment relationship in a future climate some caution is called for when using these models for extrapolation.

We have emphasized the quality of the Bayesian spatial model with narrow PFTs, but there is an intriguing possibility opened by the global model (Fig. S8, S11, and S14) – that being the representation of vegetation without reference to PFTs (1). In this case the representation of vegetation would rely entirely on the structure of trait distributions at various landscape scales (1). Such a representation eliminates the need to separately model the future locations of PFTs (or species) when inferring the future distribution of traits; hence, the output of a model like that developed here could be updated with future environmental covariates, with the caveats that 'out of sample prediction' may entail. At the same time, this method would allow for greater functional diversity than multiple PFTs with single trait values, as is currently used in most ESMs. Adopting this approach does, however, raise the issue of how to deal with the paucity of surface observations in some regions, as evidenced by the greater errors associated with estimating out of sample values with this model (Table 1). Complementary work has retrieved leaf trait maps from a global carbon cycle model fused with Earth observations (37), providing another method that could be used for direct comparison against the trait maps produced here. While the methodology outlined in our analysis brings the possibility of a PFT-free land surface closer, the current level of observations remains insufficient to provide superior prediction of trait values for all surface regions; given this, there is a necessity to continue use of PFTs for trait prediction until the paucity of trait data is rectified.

Conclusions

SLA and N_m are essential inputs into the land surface components of Earth System Models, and while phosphorus has not yet been as widely incorporated into ESMs, it has been shown

DRAFT

- particularly across the tropics - to be important to photosynthesis (9, 11, 38–41) and respiration (11, 12, 35). The maps and trait-environment relationships presented here may be used by existing land surface models that use similar categories to classify vegetation. However, it should be noted that PFT-dependent models often have many other parameters that have been calibrated to historical estimates of particular trait values (4). Thus, the values developed here, while likely drawing from a larger pool of measurements than has been done previously can not necessarily be adopted without further modification of other model elements (36, 42). Nonetheless, these results can be incorporated into a wide class of models with relative ease. We can now provide global trait distributions at the pixel scale.

The global land surface is perhaps the most heterogeneous component of the Earth System. Reducing vegetation to a collection of PFTs with fixed trait values has been the preferred method to constrain this heterogeneity and group similar biochemical and biophysical properties; however, this has been at the expense of functional diversity. This analysis quantifies the substantial magnitude of this ignored trait variation. The approach and methods presented here retain the simplicity of the PFT representation, but capture a wider range of functional diversity.

Materials and Methods

Data. The TRY database (www.try-db.org) (14) provided all data for leaf traits and the categorical traits to aggregate PFTs (TRY – Categorical Traits Dataset, https://www.try-db.org/TryWeb/Data.php#3, January 2016) used in the analysis. See Appendix 1 for a complete list of the original publications associated with this subset of TRY. The extract from TRY used here has just under 45,000 measurements of individuals from 3,680 species with measurements of at least one of specific leaf area (SLA), leaf nitrogen per dry leaf mass (N_m), and/or leaf phosphorus per leaf dry mass (P_m). The number of individual measurements varies from 32,315 for SLA on 2,953 species to 19,282 for N_m on 3,053 species down to 8,052 for P_m on 1,810 species; see Table S4 for the number of unique measurements and species found in all categorizations used in the analysis. The species taxonomy was standardized using The Plant List (43). Measurements were associated with environmental categories through Köppen-Geiger climate zones (44). All environmental variables are on a $0.5^\circ \times 0.5^\circ$ grid. Climate variables use 30 year climatologies from 1961-1990 as estimated by the Climate Research Unit (45, 46). Soil variables are from the International Soil Reference and Information Center - World Inventory of Soil Emission Potentials (ISRIC-WISE) (47). The spatial extent of PFTs have been previously estimated through satellite estimates of land cover around the year 2005 (48), and these estimates have been refined into climatic categories (15, 34). While TRY, and thus the data used here, represents the largest collection of plant traits in the world most of the measurements come from a subset of global regions: North America, Europe, Australia, China, Japan, and Brazil. There are still large sections of the planet with extremely sparse measurements, notably: much of the tropics outside of the Americas, large swathes of Central Asia, the Russian Federation, South Asia and much of the Arctic (fig. S17). Improving data collection in these regions will greatly improve future modeling efforts.

Classification of PFTs and Categorical Model. We used three nested levels of PFT classification. In the first level, all plants are categorized into a single global group ('global'). In the second level ('broad'), all plants are categorized into PFTs based on categorical traits associated with growth form (grass, shrub, tree) and leaf type (broad and needle-leaved) leading to the following four PFTs: grasses, shrubs, broad-leaved trees and needle-leaved trees (Fig. S1). In the third level ('narrow'), the broad PFTs are further refined by their climatic region – tropical, temperate, boreal – as well as leaf phenology, and, for the grasses, photosynthetic pathway (C_3 or C_4). This produces 14 PFTs (Fig. S2), which correspond exactly to those found in the community land model (CLM) (4). Note that these PFT classifications exclude non-woody eudicots ('herbs'), which were excluded from the analysis, on account of their lack of dominance within these PFT categories (49) and therefore, on account of being widely measured could overly influence the structure of the trait distributions if they were included. Satellite estimates of the PFT abundance that correspond to the "narrow" PFT categories defined above have already been calculated (15, 48) and we used these to assign a percentage of each $0.5^\circ \times 0.5^\circ$ pixel to each PFT present according to the fraction of the land surface within that pixel occupied by the PFT. The "broad" PFT fractions are calculated by summing the narrow PFT categories within each "broad" classification.

The categorical model uses the PFT categories and averages trait values for each species across individual measurements at each measured location. This defines the PFT as the interspecies range of trait values and ignores all local environmental factors. The results of the categorical model are summarized by the mean and standard deviation of each PFTs trait values (Table S4) for all three resolutions of the model. Note that in the global case where no PFT information is used, the categorical model produces a constant trait distribution across the entire vegetated world. The categorical model, and the Bayesian models described in the following section all use location specific species mean values to estimate trait distributions. We assume no intra-specific variation in trait values. However, in regions dominated by a small number of species this may lead to biased predictions. The hyper dominance of a small group of species in the Amazon has recently been demonstrated (50) and thus serves as a case study to evaluate our assumption of equal species weighting (S8, fig. S23). We found that equal weights (species means) produced trait distribution estimates closest to those of the hyper dominant trait abundances and this reinforces the use of this assumption globally. Further, as noted above, the omission of herbaceous species from tropical regions in this analysis (and (50)) may unduly limit trait diversity, and calls for further research.

Bayesian Models. A more fine-tuned depiction of geographical or spatial variation of plant trait values within each PFT can be achieved by leveraging environmental and location information, which allows trait values to adjust based on local conditions. Data for 17 climate (45, 46) and soil based (47) environmental predictors were available at the $0.5^\circ \times 0.5^\circ$ pixel resolution used to create the trait maps. To avoid overfitting and collinearity issues, these seventeen predictors were screened (see Section S7) based on correlations amongst predictors, their individual correlation with the traits, and to include climate covariates along different axes of environmental stress and both chemical and physical soil covariates. We finally selected five predictors – mean annual temperature [MAT], total annual radiation [RAD], moisture index (precipitation/evapotranspiration) [MI], percent hydrogen (aqueous) [pH], and percent clay content [CLY]. Remote sensing data products, such as Normalized Difference Vegetation Index (51)), are not used as covariates, to allow for inference outside of the historical observation period through perturbations of environmental covariates.

We utilized environment-trait relationships to obtain predictions of trait values (1, 16–18, 36, 42) in a linear regression setup. The formal details of the initial model are as follows. We denote log-transformed trait values at a geographical location s as $y_{trait}(s)$. This set of five predictors at a location s is denoted by the vector $x(s) = (x_1(s), x_2(s), ..., x_5(s))'$. A linear regression model relating the trait to the environmental predictors is specified as:

$$y_{trait}(s) = b_0 + b_1 x_1(s) + b_2 x_2(s) + ... + b_5 x_5(s) + \epsilon(s) \quad [1]$$

where b_i are the regression coefficients and $\epsilon(s)$ is the error term explaining residual variation. Estimation of model parameters and prediction were achieved with a fully Bayesian hierarchical model. This enables inclusion of prior information and prediction of full trait distributions instead of representative values (like mean or median) thereby ensuring that the uncertainty associated with the estimation

DRAFT

of model parameters is fully propagated into the predictive trait distributions.

We then generalized the above model into a Bayesian spatial linear regression model that borrows information from geographically proximal regions to capture residual spatial patterns beyond what is explained by environmental predictors. A customary specification of a spatial regression model is obtained by splitting up the error term $\epsilon(s)$ in Equation (1) into the sum of a spatial process w(s) and an error term $\eta(s)$, that accounts for the residual variation after adjusting for the spatial effects w(s). The underlying latent process w(s) accounts for local nuances beyond what is captured by the environmental predictors and is often interpreted as the net contribution from unobserved or unusable predictors. Gaussian Processes (GP) are widely used for modeling unknown spatial surfaces such as w(s), due to their convenient formulation as a multivariate Gaussian prior for the spatial random effect, unparalleled predictive performance (52) and ease of generating uncertainty quantified predictions at unobserved locations. We use the computationally effective Nearest Neighbor Gaussian Process (26) which nicely embeds into the Bayesian hierarchical setup as a prior for w(s) in the second stage of the model specification. All technical specifications of the Bayesian spatial model are provided in Section S1 of the supplementary materials.

The linear regression models used in previous studies (1, 16–18) and both the spatial and non-spatial Bayesian models described above assume a global relationship between the traits and environment. Given the goal of predicting trait values for the entire land surface, the assumption of a universal trait-environment relationship may be an oversimplification (53). Moreover, if there is significant variation in plant trait values among different PFTs, the estimated parameters will be skewed towards values from abundantly sampled PFTs, such as broad-leaved trees. Additional information about plant characteristics at a specific location, if available, can potentially be used to improve predictions. As mentioned earlier, we have PFT classifications for each observation of the dataset used here and satellite estimates of PFT abundance at all pixels. The global regression approaches described above ignores this information and can yield biased predictions at locations dominated by PFTs poorly represented in the data, such as shrubs. Hence, we also incorporate the PFT information in these regression models by allowing the trait-environment relationship to vary between different PFTs. Finally, the PFT specific distributions from the Bayesian models were weighted by the satellite based PFT abundances to create a landscape scale trait distribution, thereby enabling straightforward comparison between all three categorizations of PFT. Details of the PFT based Bayesian models are provided in Section S2. The use of a Gaussian Process based spatial model as well as the Bayesian implementation of the regression models were novel to this application of plant trait mapping and, as results indicated, were critical to improving model predictions as well as properly quantifying trait distributions.

All the code and public data are available from the authors upon request. The TRY data may be requested from the TRY database custodians.

ACKNOWLEDGMENTS. E.E.B., H.F.M., M.C., K.R.W., and P.B.R. acknowledge funding from the United States Department of Energy, Office of Science (DE-SC0012677). O.K.A. acknowledges the support of the Australian Research Council (CE140100008). P.B.R also acknowledges support from two University of Minnesota Institute on the Environment Discovery Grants. The study has been supported by the TRY initiative on plant traits (http://www.try-db.org). The TRY initiative and database is hosted, developed and maintained at the Max Planck Institute for Biogeochemistry, Jena, Germany. TRY is currently supported by DIVERSITAS/Future Earth and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. BB acknowledges a NERC independent research fellowship NE/M019160/1. JP would like to acknowledge the financial support from the European Research Council Synergy grant ERC-SyG-2013-610028 IMBALANCE-P, the Spanish Government grant CGL2013-48074-P and the Catalan Government grant SGR 2014-274. B.B.-L. was supported by the Earth System Modeling program of the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. KK acknowledges the contribution of the WUR Investment theme Resilience for the project Resilient Forest (KB-29-009-003). PM acknowledges support from ARC grant FT110100457 and NERC NE/F002149/1. WH acknowledges support from the National Natural Science Foundation of China (#41473068) and "Light of West China" Program of Chinese Academy of Sciences.

1. Van Bodegom PM, Douma JC, Verheijen LM (2014) A fully traits-based approach to modeling global vegetation distribution. *Proceedings of the National Academy of Sciences of the United States of America* 111(38):13733–8.
2. Maire V, et al. (2015) Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecology and Biogeography* 24(6):706–717.
3. DeFries RS, et al. (1995) Mapping the land surface for global atmosphere-biosphere models: Toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research* 100(D10):20867.
4. Bonan GB, et al. (2011) Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research* 116(G2):1–22.
5. Reich PB, Ellsworth DS, Walters MB (1998) Leaf structure (specific leaf area) modulates photosynthesis – nitrogen relations : evidence from within and across species and functional groups. *Functional Ecology* 12:948–958.
6. Kattge J, Knorr W, Raddatz T, Wirth C (2009) Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Global Change Biology* 15(4):976–991.
7. Crous KY, et al. (2017) Nitrogen and phosphorus availabilities interact to modulate leaf trait scaling relationships across six plant functional types in a controlled-environment study. *New Phytologist*.
8. Wright IJ, et al. (2004) The worldwide leaf economics spectrum. *Nature* 428(6985):821–827.
9. Reich PB, Oleksyn J, Wright IJ (2009) Leaf phosphorus influences the photosynthesis-nitrogen relation: A cross-biome analysis of 314 species. *Oecologia* 160(2):207–212.
10. Atkin OK, et al. (2015) Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *New Phytologist* 206(2):614–636.
11. Bahar N, et al. (2016) Leaf-level photosynthetic capacity in lowland Amazonian and high elevation, Andean tropical moist forests of Peru. *New Phytologist*.
12. Rowland L, et al. (2016) Scaling leaf respiration with nitrogen and phosphorus in tropical forests across two continents.
13. Reich PB (2005) Global biogeography of plant chemistry: Filling in the blanks. *New Phytologist* 168(2):263–266.
14. Kattge J, et al. (2011) TRY - a global database of plant traits. *Global Change Biology* 17(9):2905–2935.
15. Oleson KW, et al. (2013) Technical Description of version 4.5 of the Community Land Model (CLM), Technical Report Climate and Global Dynamics Division.
16. Reich PB, Oleksyn J (2004) Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of National Academy of Sciences* 101(30):11001–11006.
17. Ordoñez JC, et al. (2009) A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Global Ecology and Biogeography* 18(2):137–149.
18. Simpson AH, Richardson SJ, Laughlin DC (2016) Soil-climate interactions explain variation in foliar, stem, root and reproductive traits across temperate forests. *Global Ecology and Biogeography* 25(8):964–978.
19. Reich PB, Wright IJ, Lusk CH (2007) Predicting Leaf Physiology from Simple Plant and Climate Attributes : A Global GLOPNET Analysis. *Ecological Applications* 17(7):1982–1988.
20. Swenson NG, et al. (2012) The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography* 21(8):798–808.
21. Hawkins BA, Rueda M, Rangel TF, Field R, Diniz-Filho JAF (2014) Community phylogenetics at the biogeographical scale: Cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography* 41(1):23–38.
22. Šímová I, et al. (2015) Shifts in trait means and variances in North American tree assemblages: Species richness patterns are loosely related to the functional space. *Ecography* 38(7):649–658.
23. Swenson NG, et al. (2017) Phylogeny and the prediction of tree functional diversity across novel continental settings. *Global Ecology and Biogeography* pp. 1–12.
24. Douma JC, de Haan MWA, Aerts R, Witte JPM, van Bodegom PM (2012) Succession-induced trait shifts across a wide range of NW European ecosystems are driven by light and modulated by initial abiotic conditions. *Journal of Ecology* 100(2):366–380.
25. Asner GP, Knapp DE, Anderson CB, Martin RE, Vaughn N (2016) Large-scale climatic and geophysical controls on the leaf economics spectrum. *Proceedings of National Academy of Sciences*.
26. Datta A, Banerjee S, Finley A, Gelfand A (2016) Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association* 111(514):800–812.
27. Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta* 149(1):78–90.
28. Scheiter S, Higgins SI (2009) Impacts of climate change on the vegetation of Africa: An adaptive dynamic vegetation modelling approach. *Global Change Biology* 15(9):2224–2246.
29. Scheiter S, Langan L, Higgins SI (2013) Next-generation dynamic global vegetation models: learning from community ecology. *The New phytologist* 198(3):957–69.
30. Pavlick R, Drewry DT, Bohn K, Reu B, Kleidon a (2012) The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional trade-offs. *Biogeosciences Discussions* 9(4):4627–4726.
31. Pappas C, Fatichi S, Burlando P (2014) Terrestrial water and carbon fluxes across climatic gradients: does plant diversity matter? *New Phytologist* 16(i):3663.
32. Gross N, et al. (2017) Functional trait diversity maximizes ecosystem multifunctionality. *Nature Ecology & Evolution* 1(5):0132.

DRAFT

33. Clark DB, et al. (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development* 4(3):701–722.
34. Bonan GB (2002) Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles* 16(2):5.1–5.18.
35. Meir P, Grace J, Miranda AC (2001) Leaf respiration in two tropical rainforests: Constraints on physiology by phosphorus, nitrogen and temperature. *Functional Ecology* 15(3):378–387.
36. Verheijen LM, et al. (2015) Inclusion of ecologically based trait variation in plant functional types reduces the projected land carbon sink in an earth system model. *Global Change Biology* 21(8):3074–3086.
37. Bloom AA, Exbrayat JF, van der Velde IR, Feng L, Williams M (2016) The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences* (22):1–6.
38. Meir P, Levy PE, Grace J, Jarvis PG (2007) Photosynthetic parameters from two contrasting woody vegetation types in West Africa. *Plant Ecology* 192(2):277–287.
39. Domingues TF, et al. (2010) Co-limitation of photosynthetic capacity by nitrogen and phosphorus in West Africa woodlands. *Plant, Cell & Environment* 33(6):959–980.
40. Zhang Q, Wang YP, Pitman AJ, Dai YJ (2011) Limitations of nitrogen and phosphorous on the terrestrial carbon uptake in the 20th century. *Geophysical Research Letters* 38(22):1–5.
41. Medlyn B, et al. (2016) Using models to guide experiments: a priori predictions for the CO2 response of a nutrient- and water-limited mature Eucalypt woodland. *Global Change Biology* pp. 2834–2851.
42. Verheijen LM, et al. (2013) Impacts of trait variation through observed trait-climate relationships on performance of an Earth system model: A conceptual analysis. *Biogeosciences* 10(8):5497–5515.
43. 1.1 TPLV (2013) January. *http://www.theplantlist.org/.*
44. Peel B, Finlayson BL, McMahon Ta (2007) Updated world map of the Köppen-Geiger climate classification.pdf. *Hydrology and Earth System Sciences* 11:1633–1644.
45. New M, Hulme M, Jones P (1999) Representing Twentieth-Century Space – Time Climate Variability . Part I : Development of a 1961 – 90 Mean Monthly Terrestrial Climatology. *Journal of Climate* 12:829–856.
46. Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* 34(3):623–642.
47. Batjes N (2005) ISRIC-WISE: Global data set of derived soil properties on a 0.5 by 0.5 degree grid (Version 3.0). (December):1–64.
48. Lawrence PJ, Chase TN (2007) Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *Journal of Geophysical Research: Biogeosciences* 112(1).
49. Gibson DJ (2009) *Grasses & Grassland Ecology*. (Oxford University Press, New York).
50. ter Steege H, et al. (2013) Hyperdominance in the Amazonian tree flora. *Science* 342(6156):1243092.
51. Ollinger SV, et al. (2008) Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences of the United States of America* 105(49):19336–41.
52. Rasmussen C (1996) *Evaluation of Gaussian Processes and other methods for non-linear regression*. (University of Toronto), Phd thesis edition.
53. Verheijen LM, Aerts R, Bönisch G, Kattge J, Van Bodegom PM (2016) Variation in trait tradeoffs allows differentiation among predefined plant functional types: Implications for predictive ecology. *New Phytologist* 209(2):563–575.

DRAFT

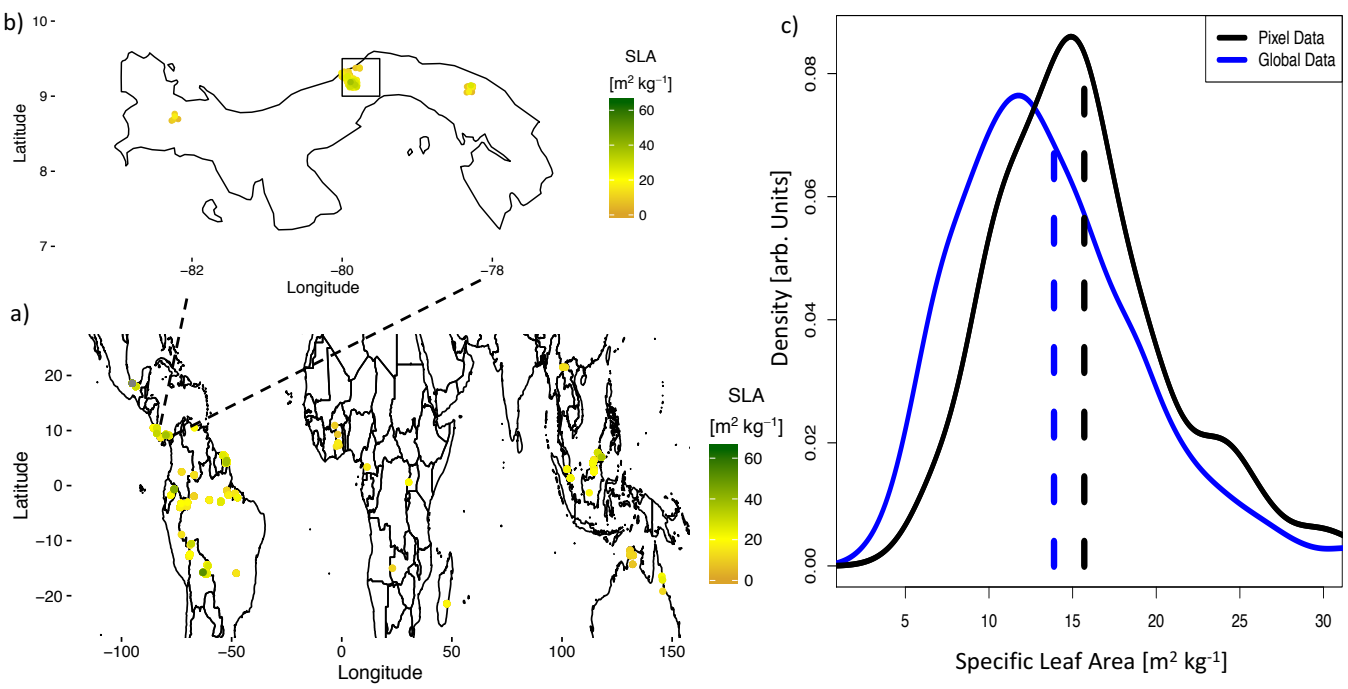

Fig. 1. Trait data a) Global locations and values of specific leaf area measurements for the PFT Tropical Broadleaf Evergreen Trees. b) Locations and values of specific leaf area measurements for the Tropical Broadleaf Evergreen Trees in Panama. The central square indicates a $0.5^\circ \times 0.5^\circ$ pixel containing the Barro Colorado Island sites (see Fig. 5). These points have been jittered up to 0.05° to highlight the density of measurements. c) The full distribution of specific leaf area values for all species classified as the Evergreen Broadleaf Tropical Trees. The blue line is the global data while black is the local pixel, the dashed vertical lines are the respective means.

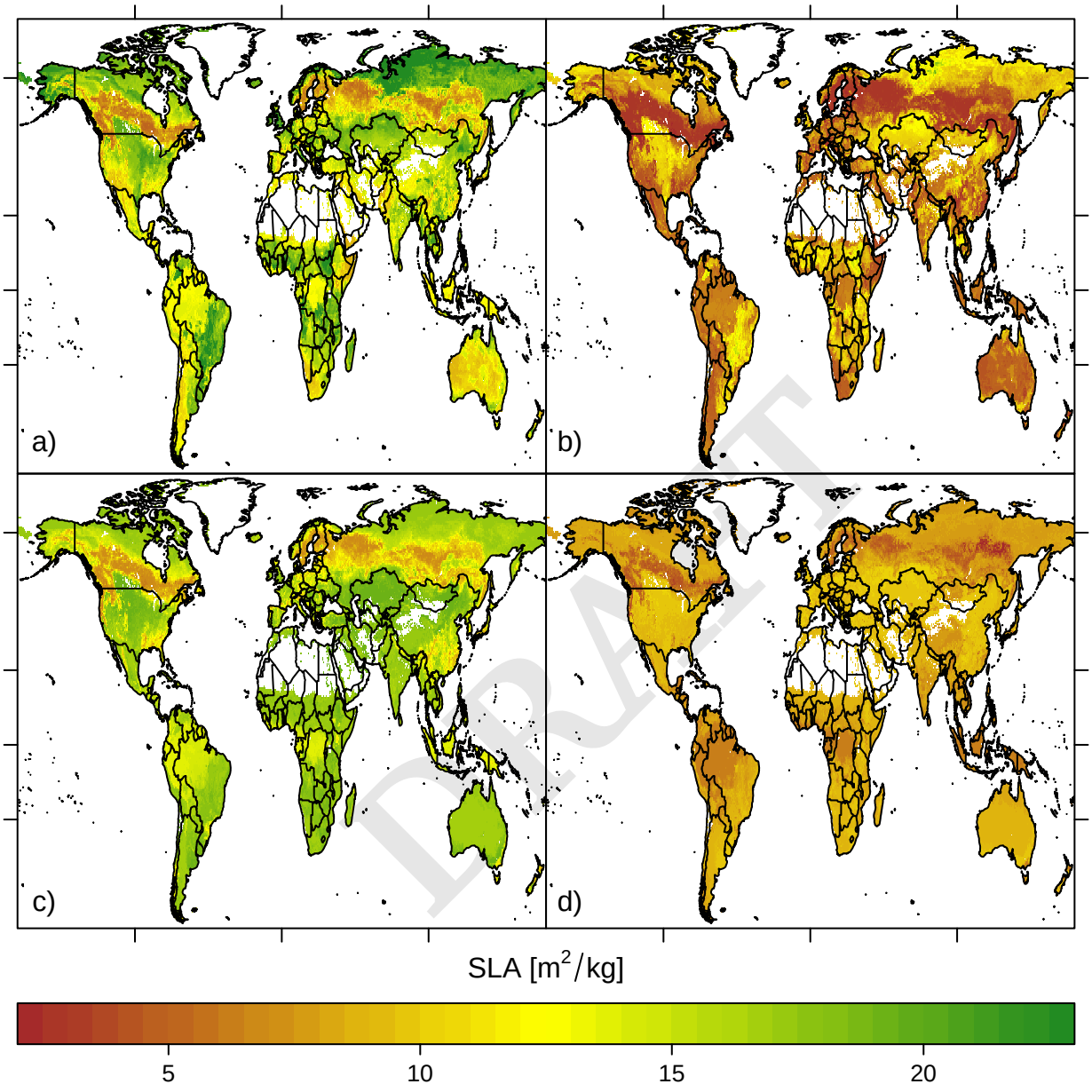

Fig. 2. Specific Leaf Area maps a) Bayesian spatial model pixel mean estimates b) Bayesian spatial model pixel standard deviation estimates c) Categorical model pixel mean estimates d) Categorical model pixel standard deviation estimates. For clarity, the color bars have been truncated at the compound 5th and 95th percentiles of both models. Latitude tick marks indicate the equator, tropics, and arctic circle and longitude is marked at -100°E, 0°, and 100°E.

DRAFT

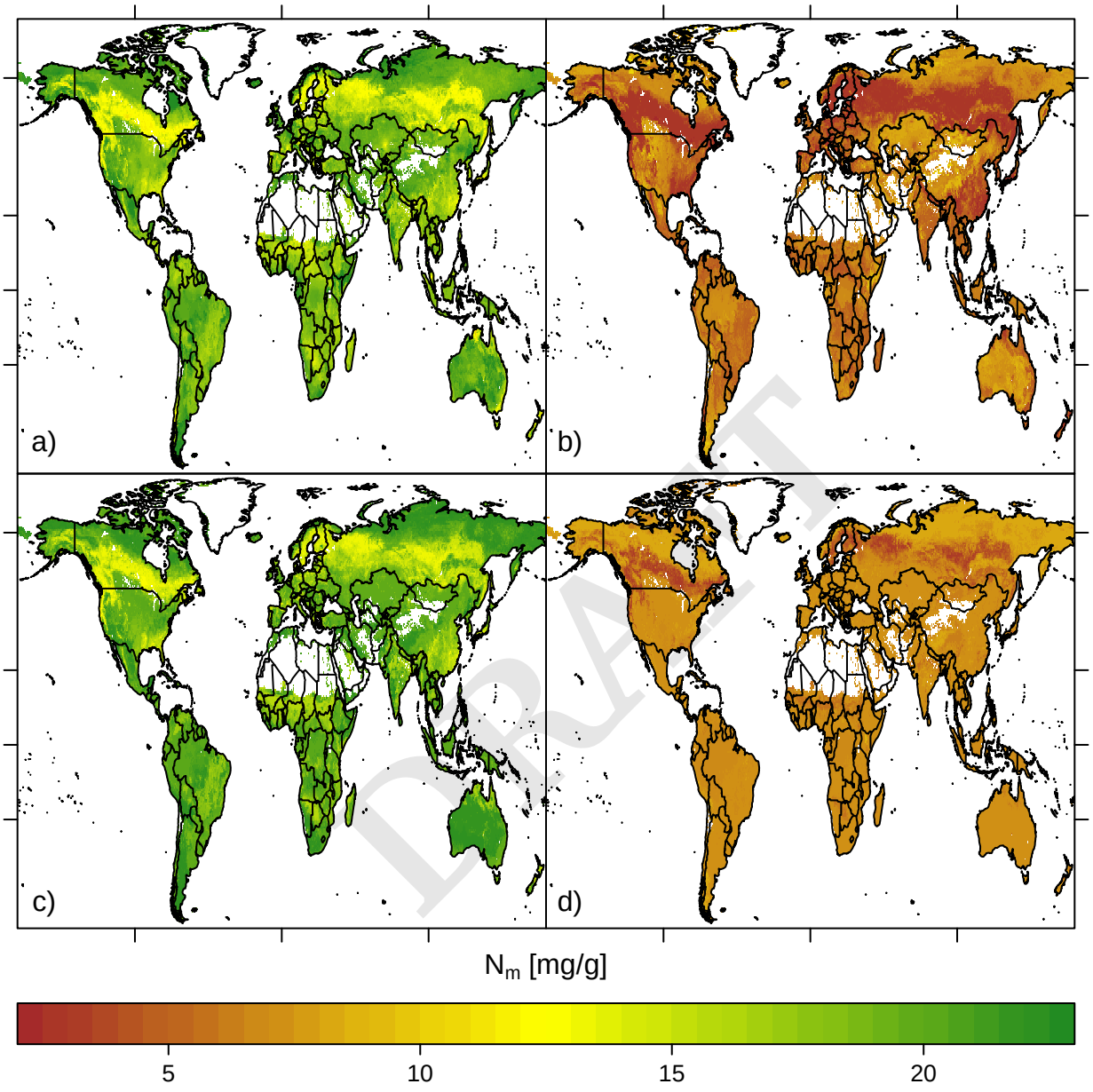

Fig. 3. Nitrogen [mass] maps a) Bayesian spatial model pixel mean estimates b) Bayesian spatial model pixel standard deviation estimates c) Categorical model pixel mean estimates d) Categorical model pixel standard deviation estimates. For clarity, the color bars have been truncated at the compound 5th and 95th percentiles of both models. Latitude tick marks indicate the equator, tropics, and arctic circle and longitude is marked at -100°E, 0°, and 100°E.

DRAFT

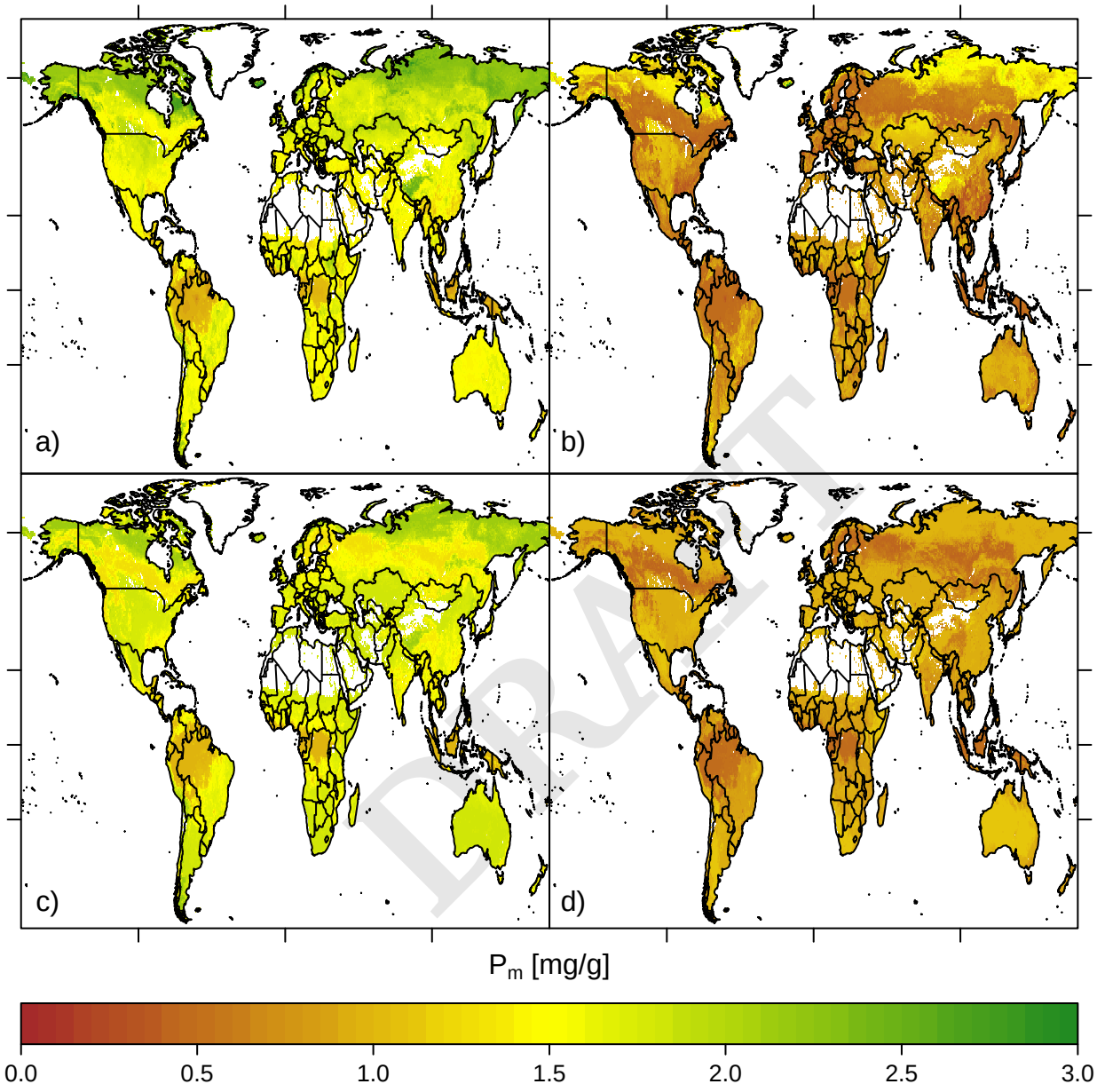

Fig. 4. Phosphorus [mass] maps a) Bayesian spatial model pixel mean estimates b) Bayesian spatial model pixel standard deviation estimates c) Categorical model pixel mean estimates d) Categorical model pixel standard deviation estimates. For clarity, the color bars have been truncated at the compound 5th and 95th percentiles of both models. Latitude tick marks indicate the equator, tropics, and arctic circle and longitude is marked at -100°E, 0°, and 100°E.

DRAFT

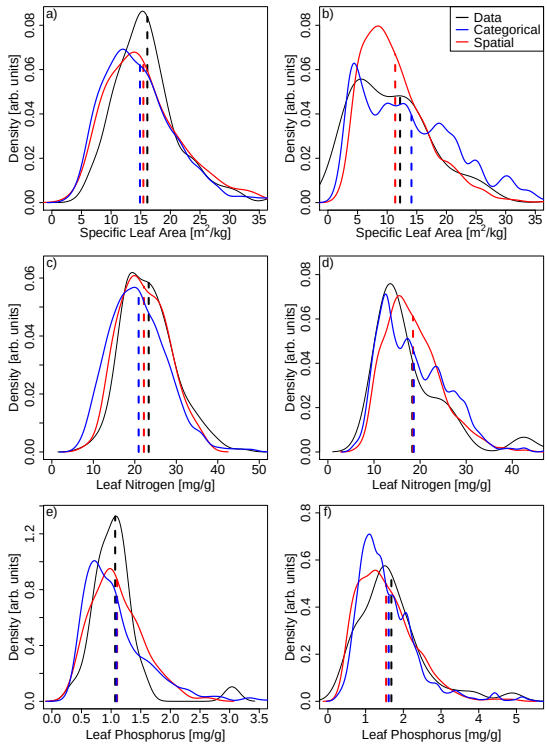

Fig. 5. Empirical trait distributions Barro Colorado Island on the left (a, c, e) and the US Southwest on the right (b, d, f). The first row is SLA (a, b), the second is leaf nitrogen (c, d) and the third is leaf phosphorus (e, f). Each panel depicts the distribution of the data in solid black, the categorical model in blue and the Bayesian spatial model in red. The vertical lines indicate mean values.

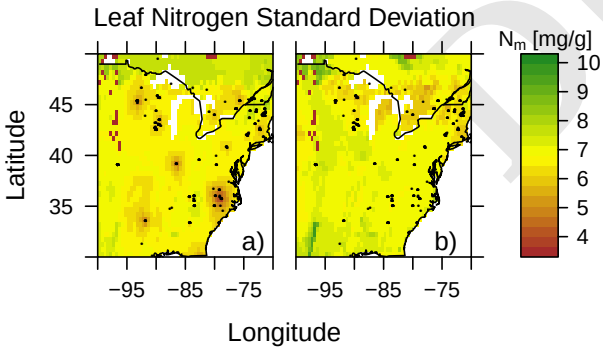

Fig. 6. Spatial learning a) the spatial model standard deviation of Nm. The predicted variation near the data locations (black dots) are much lower than variation at locations away from any data point. b) the linear model standard deviation which does not account for local spatial information has no such pattern.

Table 1. Model evaluation

SLA			
Model	ps-R^2	RMSPE	CP
Cg	NA	8.13	91.2%
Cb	16.9%	7.13	94.7%
Cn	26.0%	6.66	95.8%
Lg	4.6%	7.99	91.3%
Lb	23.4%	6.93	94.0%
Ln	30.7%	6.53	**95.2%**
Sg	45.5%	7.54	93.6%
Sb	58.5%	6.31	97.7%
Sn	**60.2%**	**6.13**	97.7%
N_m			
Model	ps-R^2	RMSPE	CP
Cg	NA	7.16	93.3%
Cb	12.5%	6.95	93.2%
Cn	19.4%	6.47	92.7%
Lg	5.2%	7.28	93.2%
Lb	16.7%	6.71	94.3%
Ln	24.1%	6.42	**94.6%**
Sg	44.2%	7.19	93.6%
Sb	53.7%	6.36	96.1%
Sn	**54.8%**	**6.18**	96.1%
P_m			
Model	ps-R^2	RMSPE	CP
Cg	NA	0.86	90.5%
Cb	5.3%	0.86	90.5%
Cn	28.1%	**0.78**	91.1%
Lg	25.6%	0.84	87.2%
Lb	32.8%	0.85	85.3%
Ln	35.4%	0.82	87.0%
Sg	62.0%	0.83	90.7%
Sb	66.7%	0.81	**92.0%**
Sn	**67.6%**	0.80	91.3%

The pseudo-R2 (ps-R2), RMSPE and CP statistics for all nine models, for each of the three traits. The bold entries correspond to the model producing highest ps-R2, lowest RMSPE, or CP closest to 0.95. The categorical global model (Cg) produces a constant estimate and hence ps-R2 is not defined. Each model is indicated by a two-letter abbreviation: C=Categorical, L=Linear, S=Spatial and the accompanying PFT resolution: g=global, b=broad, n=narrow.

DRAFT